THE PROBLEM OF SELECTING A GIVEN NUMBER OF REPRESENTATIVE
POINTS IN A NORMAL POPULATION AND A
GENERALIZED MILLS' RATIO

TECHNICAL REPORT NO. 5

KAI-TAI FANG   and   SHU-DONG HE

OCTOBER 1982

DEPARTMENT OF STATISTICS
STANFORD   UNIVERSITY
STANFORD, CALIFORNIA

THE PROBLEM OF SELECTING A GIVEN NUMBER OF REPRESENTATIVE

POINTS IN A NORMAL POPULATION AND A

GENERALIZED MILLS' RATIO

TECHNICAL REPORT NO. 5

Kai-Tai Fang
Institute of Applied Mathematics, Academia Sinica

Shu-Dong He
Huazheng Institute of Technology, Wuhan, China

OCTOBER 1982

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

THE PROBLEM OF SELECTING A GIVEN NUMBER OF REPRESENTATIVE

POINTS IN A NORMAL POPULATION AND A

GENERALIZED MILLS' RATIO

By

Kai-Tai Fang and Shu-Dong He

1.   Introduction.

The problem of selecting a given number of representative points to
retain as much information of the population as possible arises in many
situations.

For example, in order to standardize clothes, we take  p  measure-
ments  of the body of each of  n  individuals (in general, n is sufficiently
large), and project these  p  dimensional data onto a  q  dimensional
space (q = 1,2 or 3) by principal components analysis of by some other
method.  We wish to select  m  points that best represent the data in the
q-dimensional space.  In Fang (1976), this problem is analyzed for one
and two-dimensional normal distributions where the intervals are of equal
lengths and the  m  points are centered in each interval.

Bofinger (1970) studied the question of grouping a continuous bi-
variate distribution by intervals on the marginals thereby obtaining a
discrete bivariate distribution.  She sought the grouping that would
provide the maximum possible correlation between these marginal variables.
The solution is approximate when both marginals are grouped and exact when
only one margin is grouped.  For the bivariate normal distribution with
one margin grouped, tables of interval end points that maximize the
correlation are provided up to 10 intervals.

1

Prior to this, Sitgreaves (1961) arrived at the same bivariate model as Bofinger in connection with a query about determining the optimal item difficulties in a special mental test design in psychometrics. The optimal representative points, for k = 2,3,4,5 intervals under a bivariate normal structure are determined in that paper by a graphical procedure thus producing the optimal item difficulties for a test with k items.

A paper by Max (1960) seeks to quantize the univariate normal distribution so that it can be represented by k points. Employing a mean square error loss function, Max provides tables for k = 1,2,...,36 to yield optimal representative points for digitizing the univariate normal. There is a previous paper by Cox (1957) whose motivation is quantization and who for the univariate normal provides a table of optimal representative points for k = 2,...,6 groups. In Anderberg (1973) there is discussion of sectionalizing the univariate normal distribution to transform interval data to ordinal data for clustering procedures and some abridged tables are given.

Max was motivated by a signal processing problem and it is in this vein that Zador (1963) considered how to select a random discrete vector in one and higher dimensions to approximate a continuous variable employing a mean square error loss function. He provided a generalized model for the multivariate normal distribution and secured bounds on quantization error as a function of dimension and moments of the error term as the number of representative points increased. A revised version of this work appears in Zador (1982). The tables produced by these investigators and the tables in this report yield the same listings except for computational accuracy. However, this report and the paper by Max contain the most extensive tables.

We have been interested in this subject since we worked on the standardization of clothes. In October 1981 we, just as previous authors, obtained independently the results in this paper. Later Professor T. W. Anderson and V. Srinivasan told us of Cox's and Bofinger's work. Recently, Professor H. Solomon introduced Zador, Max and Sitgreave's work to us. It is a very interesting story in the history of Statistics that several investigators, motivated by different applications were led in the same methods, and obtained independently similar results over a period of more than twenty years. Compared to other papers, ours gives more theoretical proof of the computation of the representative points. Perhaps it is valuable for people who want to compute more representative points or to compute the representative points in two dimensional space. In addition we give some basic results on the generalized Mills' ratio that may be useful in statistical analysis.

Suppose the distribution of the population is $N(\mu, \sigma^2)$, where $\mu$ and $\sigma$ are known. We wish to determine $m$ points $x_1, x_2, \ldots, x_m$ that are the best representatives of the population.

What is the meaning of "best"? We consider the loss function

$$(1.1) \qquad f(x_1, \ldots, x_m) = \int_{-\infty}^{\infty} \min_{1 \leq i \leq m} \left(\frac{x_i - x}{\sigma}\right)^2 \phi(x) \, dx$$

where $\phi(x)$ denotes the density of the normal variate with mean $\mu$ and variance $\sigma^2$. Without loss of generality, we can assume $\mu = 0$, and $\sigma = 1$.

In order to study properties of the solutions, we generalized Mills' ratio, and give some basic properties of the generalized Mills' ratio in Section 3. In Section 4 we discuss properties of the solution of some equations. As a consequence, a computational procedure is suggested and a table of $x_1, \ldots, x_m$ for $m \leq 31$ obtained by computer is given in Section 5.

3

## 2. Preliminaries.

Rewrite (1.1) with $\mu = 0$, $\sigma = 1$ as

$$(2.1) \qquad f(x_1,\ldots,x_m) = \int_{-\infty}^{(x_1+x_2)/2} (x-x_1)^2 \phi(x)\,dx$$

$$+ \int_{(x_1+x_2)/2}^{(x_2+x_3)/2} (x-x_2)^2 \phi(x)\,dx + \cdots + \int_{(x_{m-1}+x_m)/2}^{+\infty} (x-x_m)^2 \phi(x)\,dx \;,$$

where $x_1 < \cdots < x_m$ and $\phi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$. In order to find $x_1,\ldots,x_m$, we solve the first derivative equations:

$$\frac{\partial f(x_1,\ldots,x_m)}{\partial x_i} = 0 \;, \qquad i = 1,2,\ldots,m \;.$$

We obtain the following equations:

$$(2.2) \qquad \begin{cases} \displaystyle\int_{-\infty}^{\frac{1}{2}(x_1+x_2)} (x-x_1)\phi(x)\,dx = 0 \;, \\[2em] \displaystyle\int_{\frac{1}{2}(x_1+x_2)}^{\frac{1}{2}(x_2+x_3)} (x-x_2)\phi(x)\,dx = 0 \;, \\[2em] \qquad \cdots \qquad\qquad \cdots \\[2em] \displaystyle\int_{\frac{1}{2}(x_{m-1}+x_m)}^{+\infty} (x-x_m)\phi(x)\,dx = 0 \;. \end{cases}$$

**Lemma 1.** The solution of the equations (2.2) is symmetric about the origin, i.e., $x_1 = -x_{m-i+1}$, $i=1,2,\ldots,m$.

4

Proof. Let $X \sim N(0,1)$, using the symmetry of the density of $X$ about the origin we have

$$E(\min(x_i-x)^2) = E(\min(x_i+x)^2) = E(\min(-x_i-x)^2)$$

the lemma follows.    Q.E.D.


From Lemma 1, we only find $0 < x_1 < x_2 < \cdots < x_k$ if $m = 2k$ is even and $0 = x_0 < x_1 < \cdots < x_k$ if $m = 2k+1$ is odd. And the loss functions become respectively

$$(2.3) \qquad f(x_1,\ldots,x_m) = \int_0^{\frac{1}{2}(x_1+x_2)} (x-x_1)^2\phi(x)dx$$

$$+ \int_{\frac{1}{2}(x_1+x_2)}^{\frac{1}{2}(x_2+x_3)} (x-x_2)^2\phi(x)dx + \cdots + \int_{\frac{1}{2}(x_{k-1}+x_k)}^{\infty} (x-x_k)^2\phi(x)dx \ ,$$

$$(2.4) \qquad f(x_0,\ldots,x_m) = \int_0^{\frac{1}{2}x_1} x^2\phi(x)dx$$

$$+ \int_{\frac{1}{2}x_1}^{\frac{1}{2}(x_1+x_2)} (x-x_1)^2\phi(x)dx + \cdots + \int_{\frac{1}{2}(x_{k-1}+x_k)}^{\infty} (x-x_k)^2\phi(x)dx \ .$$


Let $\partial f(x_1,\ldots,x_k)/\partial x_i = 0$ and $\partial f(x_0,x_1,\ldots,x_k)/\partial x_i = 0$, $i = 1,2,\ldots,k$, and we obtain the following two systems of equations:    if $m = 2k$

$$(2.5) \begin{cases} \phi(0) - \phi(\tfrac{1}{2}(x_1+x_2)) = x_1[\Phi(\tfrac{1}{2}(x_1+x_2)) - \Phi(0)] \ , \\[2mm] \phi(\tfrac{1}{2}(x_1+x_2)) - \phi(\tfrac{1}{2}(x_2+x_3)) = x_2[\Phi(\tfrac{1}{2}(x_2+x_3)) - \Phi(\tfrac{1}{2}(x_1+x_2))] \ , \\[1mm] \qquad \cdots \qquad\qquad \cdots \qquad\qquad \cdots \\[1mm] \phi(\tfrac{1}{2}(x_{k-2}+x_{k-1})) - \phi(\tfrac{1}{2}(x_{k-1}+x_k)) = x_{k-1}[\Phi(\tfrac{1}{2}(x_{k-1}+x_k)) - \Phi(\tfrac{1}{2}(x_{k-1}+x_{k-2}))], \\[2mm] \phi(\tfrac{1}{2}(x_{k-1}+x_k)) = x_k[1 - \Phi(\tfrac{1}{2}(x_{k-1}+x_k))] \ , \end{cases}$$

and if $m = 2k+1$,

$$(2.6) \begin{cases} \phi(\tfrac{1}{2} x_1) - \phi(\tfrac{1}{2}(x_1+x_2)) = x_1[\Phi(\tfrac{1}{2}(x_1+x_2)) - \Phi(\tfrac{1}{2} x_1)] \ , \\[2mm] \phi(\tfrac{1}{2}(x_1+x_2)) - \phi(\tfrac{1}{2}(x_2+x_3)) = x_2[\Phi(\tfrac{1}{2}(x_2+x_3)) - \Phi(\tfrac{1}{2}(x_1+x_2))] \ , \\[1mm] \qquad \cdots \qquad\qquad \cdots \qquad\qquad \cdots \\[1mm] \phi(\tfrac{1}{2}(x_{k-2}+x_{k-1})) - (\tfrac{1}{2}(x_{k-1}+x_k)) = x_{k-1}[\Phi(\tfrac{1}{2}(x_{k-1}+x_k)) - \Phi(\tfrac{1}{2}(x_{k-2}+x_{k-1}))] \ , \\[2mm] \phi(\tfrac{1}{2}(x_{k-1}+x_k)) = x_k[1 - (\tfrac{1}{2}(x_{k-1}+x_k))] \ . \end{cases}$$

Where $\Phi(x)$ is the normal cumulative distribution function with mean $0$ and variance $1$. For any $0 \le x < y \le \infty$ define a function

$$(2.7) \qquad\qquad M(x,y) = \frac{\Phi(y) - \Phi(x)}{\phi(x) - \phi(y)} \ .$$

When $y = \infty$, $M(x,\infty)$ is the usual Mills' ratio $M(x)$; thus we call $M(x,y)$ the generalized Mills' ratio. Now the systems of equations (2.5) and (2.6) can be rewritten as follows with $x_{k+1} \equiv \infty$:

$$(2.8) \quad x_i M[(\tfrac{1}{2}(x_{i-1}+x_i), \ \tfrac{1}{2}(x_i+x_{i+1})] = 1, \quad i = 1,2,\dots,k \ ,$$

where $x_0 = -x_1$ if $m = 2k$ and $x_0 = 0$ if $m = 2k+1$. In order to solve the system of equations (2.8), it is necessary to study some properties of $M(x,y)$.

3. Generalized Mills' ratio.

It can be verified that the generalized Mills' ratio defined by (2.2) satisfies

$$(3.1) \qquad M_x'(x,y) = \frac{\partial}{\partial x} M(x,y) = (1-e^{-\frac{1}{2}(y^2-x^2)})^{-1}[xM(x,y)-1] \,,$$

$$(3.2) \qquad M_y'(x,y) = \frac{\partial}{\partial y} M(x,y) = (e^{\frac{1}{2}(y^2-x^2)}-1)^{-1}[1-yM(x,y)] \text{ (for } y < \infty).$$

In general, we have by induction

$$(3.3) \qquad \frac{\partial^{k+\ell}}{\partial^k x \partial^\ell y} M(x,y) = u_{k,\ell}(x,y)M(x,y)-v_{k,\ell}(x,y) \,,$$

where $u_{k,\ell}(x,y)$ and $v_{k,\ell}(x,y)$ are defined by the following recursive formulae:

$$(3.4) \qquad \begin{cases} u_{0,0}(x,y) = 1, \; v_{0,0}(x,y) = 0 \\[2mm] u_{k+1,\ell}(x,y) = \frac{\partial}{\partial x} u_{k,\ell}(x,y) + xq(x,y)u_{k,\ell}(x,y) \,, \\[2mm] v_{k+1,\ell}(x,y) = q(x,y)u_{k,\ell}(x,y) + \frac{\partial}{\partial x} v_{k,\ell}(x,y) \,, \\[2mm] u_{k,\ell+1}(x,y) = \frac{\partial}{\partial y} u_{k,\ell}(x,y) + y(1-q(x,y))u_{k,\ell}(x,y) \,, \\[2mm] v_{k,\ell+1}(x,y) = (1-q(x,y))u_{k,\ell}(x,y) + \frac{\partial}{\partial y} v_{k,\ell}(x,y) \,, \end{cases}$$

and

$$(3.5) \qquad q(x,y) = (1-e^{-\frac{1}{2}(y^2-x^2)})^{-1} \,.$$

7

In particular, from (3.4) we find that

$$u_{1,0}(x,y) = xq(x,y), \qquad\qquad v_{1,0}(x,y) = q(x,y),$$

$$u_{0,1}(x,y) = y(1-q(x,y)), \qquad\qquad v_{0,1}(x,y) = 1-q(x,y),$$

$$u_{1,1}(x,y) = 2xyq(x,y)(1-q(x,y)), \qquad v_{1,1}(x,y) = (x+y)q(x,y)(1-q(x,y)),$$

$$u_{2,0}(x,y) = 2x^2q^2(x,y)+(1-x^2)q(x,y), \quad v_{2,0}(x,y) = 2xq^2(x,y)-xq(x,y),$$

$$u_{0,2}(x,y) = 2y^2q^2(x,y)-(3y^2+1)q(x,y)+(y^2+1),$$

$$v_{0,2}(x,y) = y[1-3q(x,y)+2q^2(x,y)] \ .$$

Consequently, $u_{k,\ell}(x,y)$ and $v_{k,\ell}(x,y)$ are polynomials in $q(x,y)$, i.e

$$u_{k,\ell}(x,y) = \sum_{j=0}^{k+\ell} a_j^{(k,\ell)}(x,y)q^j(x,y) \ ,$$

(3.6)

$$v_{k,\ell}(x,y) = \sum_{j=0}^{k+\ell} b_j^{(k,\ell)}(x,y)q^j(x,y) \ ,$$

where $a_j^{(k,\ell)}(x,y)$ and $b_j^{(k,\ell)}(x,y)$ are polynomials of $x$ and $y$ in which the power of $x$ is less or equal to $k$ and the power of $v$ is less or equal to $\ell$. By using (3.4) and (3.6), we can obtain the recursive formulae for $a_j^{(k,\ell)}(x,y)$ and $b_j^{(k,\ell)}(x,y)$, too.

It is well known that (cf. Feller (1968), p. 193)

$$(3.7) \qquad M(x) \sim \frac{1}{x} - \frac{1}{x^3} + \frac{1.3}{x^5} - \frac{1.3.5}{x^7} + \cdots + (-1)^n \frac{(2n-1)!!}{x^{2n+1}} \ ,$$

where $(2n-1)!! = (2n-1)(2n-3)\cdots 3.1$. Similarly, by using integration by parts we can find

8

$$(3.8) \qquad (e^{-\frac{1}{2}x^2} - e^{-\frac{1}{2}y^2}) M(x,y) = \int_x^y e^{-\frac{t^2}{2}} dt$$

$$= (\frac{1}{x} e^{-\frac{1}{2}x^2} - \frac{1}{y} e^{-\frac{1}{2}y^2}) - \int_x^y \frac{1}{t^2} e^{-\frac{1}{2}t^2} dt$$

$$= (\frac{1}{x} e^{-\frac{1}{2}x^2} - \frac{1}{y} e^{-\frac{1}{2}y^2}) - (\frac{1}{x^3} e^{-\frac{1}{2}x^2} - \frac{1}{y^3} e^{-\frac{1}{2}y^2}) + 3 \int_x^y \frac{1}{t^4} e^{-\frac{1}{2}t^2} dt$$

$$(3.9) \qquad = \sum_{j=0}^{k} (-1)^j (2j-1)!! (x^{-(2j+1)} e^{-\frac{1}{2}x^2} - y^{-(2j+1)} e^{-\frac{1}{2}y^2})$$

$$+ (-1)^{k+1} (2k+1)!! \int_x^y t^{-2k} e^{-\frac{1}{2}t^2} dt$$

$$\equiv I_{1k} + I_{2k} \quad (say) .$$

Then $(e^{-\frac{1}{2}x^2} - e^{-\frac{1}{2}y^2})^{-1} I_{1k}$ is an overestimate of $M(x,y)$ if $k$ is zero or even and is an underestimate of $M(x,y)$ if $k$ is odd. The following lemma is useful later.

Lemma 2. $M_x'(x,y)$ and $M_y'(x,y)$ are negative for $0 \le x < y \le \infty$ and for $0 \le x < y < \infty$, respectively.

Proof. From (3.1) and (3.2) we only prove that

$$(3.10) \qquad x M(x,y)-1 < 0 \qquad \text{for} \quad 0 \le x < y \le \infty ,$$

and

$$(3.11) \qquad 1-y M(x,y) < 0 \qquad \text{for} \quad 0 \le x < y < \infty .$$

By using (3.8) we have

$$xM(x,y)-1 = (e^{-\frac{1}{2}x^2}-e^{-\frac{1}{2}y^2})^{-1}[e^{-\frac{1}{2}y^2}(1-\frac{x}{y}) - x \int_x^y \frac{1}{t^2} e^{-\frac{1}{2}t^2}dt] \ .$$

For any fixed $x \geq 0$, define a function of $y$

$$g(y) = e^{-\frac{1}{2}y^2}(1-\frac{x}{y}) - x \int_x^y \frac{1}{t^2} e^{-\frac{1}{2}t^2}dt \ , \qquad y > x \ .$$

Then $xM(x,y)-1 < 0$ if and only if $g(y) < 0$ for $y > x$. The latter is true because $g(x+0) = 0$, $g(\infty) = -x \int_x^\infty \frac{1}{t^2} e^{-\frac{1}{2}t^2}dt < 0$ and

$$g'(y) = -(y-x)e^{-\frac{1}{2}y^2} < 0 \ .$$

The inequality (3.10) can be proved by a similar method. Q.E.D.

In the next section we need the following lemma.

Lemma 3. For $x > 0$

$$\phi(x)(x+2M(x)) < 1 \ .$$

Proof. Let $f(x) = \phi(x)(x+2M(x))-1$. Then the lemma follows the fact that $f(0) = 0$ and $f'(x) = -\phi(x)(1+x^2) < 0$.

4. **Some Properties of the Equations.**

We wish to give a procedure to find the solutions of both systems of equations (2.5) and (2.6). The idea is that given a suitable $x_1$ we find $x_2$ from the first equation, then for fixed $x_1$ and $x_2$ obtain $x_3$ from the second equation, based on the $x_2$ and $x_3$ we get $x_4$ from the third equation, finally we obtain $x_k$ from the last second equation. In other words, for fixed $x_{k-1}$ we can get another solution $x_k^*$ from the

10

last equation. If the difference between $x_k$ and $x_k^*$ is very small, then $x_1, \ldots, x_k$ are the solutions required; otherwise we modify $x_k$ and repeat the above process. We will prove the process approaches the solution.

The equations in (2.5) and (2.6) can be classified into four groups. (They are (4.1), (4.2), (4.5) and (4.7).) Now we discuss their properties respectively.

Let $x_{m1}, \ldots, x_{mk}$ and $x_{m0}(=0), x_{m1}, \ldots, x_{mk}$ denote the solutions of (2.5) and (2.6), respectively. When $m = 2$, there is only one equation in (2.5), i.e.

$$\phi(0) = x_1(1-\Phi(0)) = \tfrac{1}{2}x_1 \ .$$

Thus $x_{21} = 2\phi(0) = \sqrt{2/\pi} \doteq .797885$.

<u>Theorem 1</u>. For any given $x_1 > 0$, the equation

$$(4.1) \qquad\qquad x_1 M(0, \tfrac{1}{2}(x_1+x_2)) = 1$$

or

$$(4.1)' \qquad \phi(0)-\phi(\tfrac{1}{2}(x_1+x_2)) = x_1(\Phi(\tfrac{1}{2}(x_1+x_2))-\Phi(0))$$

has a unique solution $x_2 = g_2(x_1)$, say, if and only if $x_1 < x_{21}$. If the condition is satisfied the function $g_2(x_1)$ is strictly increasing.

<u>Proof</u>. Let

$$F(x_1,x_2) = 1-x_1 M(0, \tfrac{1}{2}(x_1+x_2)) \ , \qquad x_2 \geq x_1 \ .$$

From Lemma 2 we have

$$F'_{x_2}(x_1,x_2) = -\tfrac{1}{2} x_1 M'_y(0,y)\Big|_{y=\frac{1}{2}(x_1+x_2)} > 0$$

and

$$F(x_1,x_1) = 1-x_1 M(0,x_1) < 0 \; .$$

Thus for given $x_1 > 0$ the equation (4.1) has a unique solution if and only if $F(x_1,\infty) > 0$. As

$$F(x_1,\infty) = 1-x_1 M(0,\infty) = 1-x_1 M(0) = 1-x_1 \sqrt{\pi/2} \; .$$

$F(x_1,\infty) > 0$ if and only if $x_1 < \sqrt{2/\pi} = x_{21}$.

If $x_{21} > x_1$, $g_2(x_1)$ is strictly increasing and if $dx_2/dx_1 = -G'_{x_1}(x_1,x_2)/G'_{x_2}(x_1,x_2) > 0$, where (cf. (4.1)')

$$G(x_1,x_2) = \phi(0)-\phi(\tfrac{1}{2}(x_1+x_2))-x_1(\Phi(\tfrac{1}{2}(x_1+x_2))-\tfrac{1}{2}) \; .$$

It is easily shown that

$$G'_{x_2}(x_1,x_2) = \tfrac{1}{4}(x_2-x_1)\phi(\tfrac{1}{2}(x_1+x_2)) > 0$$

and

$$G'_{x_1}(x_1,x_2) = \tfrac{1}{4}(x_2-x_1)\phi(\tfrac{1}{2}(x_1+x_2))- (\Phi(\tfrac{1}{2}(x_1+x_2))-\tfrac{1}{2})$$

$$= H(x_1,x_2) \; ,$$

say. From Lemma 3

$$H(0,x_2) = \tfrac{1}{4} x_2 \phi(\tfrac{1}{2} x_2) - \Phi(\tfrac{1}{2} x_2) + \tfrac{1}{2}$$

$$= \tfrac{1}{2}[\tfrac{1}{2} x_2 \phi(\tfrac{1}{2} x_2) + 2(1-\Phi(\tfrac{1}{2} x_2))-1]$$

$$= \tfrac{1}{2} \phi(\tfrac{1}{2} x_2)[\tfrac{1}{2} x_2 + 2M(\tfrac{1}{2} x_2)-1] < 0 .$$

As

$$H'_{x_1}(x_1,x_2) = -\tfrac{1}{4} \phi(\tfrac{1}{2}(x_1+x_2))(3+\tfrac{1}{4}(x_2^2-x_1^2)) < 0 ,$$

$H(x_1,x_2) < 0$ for $0 \leq x_1 < x_2$, thus $dx_2/dx_1 > 0$. Q.E.D.

As $x_2 = g_2(x_1)$ is a monotonic function, there exist $a = g_2(0+)$ and $b = g_2(x_{21})$. It can be verified that $a = 0$ and $b = \infty$, i.e., the function $x_2 = g_2(x_1)$ is increasing from $0$ to $\infty$ when $x_1$ is from $0$ to $x_{21}$.

<u>Theorem 2.</u> For any given $x_1 \geq 0$, the equation

(4.2)
$$x_2 M(\tfrac{1}{2}(x_1+x_2)) = 1$$

or

(4.2)'
$$\phi(\tfrac{1}{2}(x_1+x_2)) = x_2(1-\Phi(\tfrac{1}{2}(x_1+x_2)))$$

has a unique solution $x_2 = h(x_1)$ and $h(x_1)$ is strictly increasing.

<u>Proof.</u> Let

$$F(x_1,x_2) = \phi(\tfrac{1}{2}(x_1,x_2))-x_2(1-\Phi(\tfrac{1}{2}(x_1+x_2))), \quad x_2 \geq x_1 \geq 0 .$$

By (3.9) we have

13

$$F(x_1, x_1) = \phi(x_1) - x_1(1 - \Phi(x_1)) = \phi(x_1)(1 - x_1 M(x_1)) > 0$$

and $F(x_1, \infty) = 0$. Further we have

$$F'_{x_2}(x_1, x_2) = \phi(\tfrac{1}{2}(x_1 + x_2))[\tfrac{1}{4}(x_2 - x_1) - M(\tfrac{1}{2}(x_1 + x_2))] \ .$$

As $M(x)$ is strictly increasing with $M(0) = \sqrt{\pi/2}$ and $t(x_2) \equiv \tfrac{1}{4}(x_2 - x_1)$ is a strictly increasing function of $x_2$ with $t(0) = -\tfrac{1}{4} x_1$, there exists an $x_0$ (depending on $x_1$) such that $F'_{x_2}(x_1, x_2) < 0$ if $x_2 < x_0$, $F'_{x_2}(x_1, x_2) > 0$ if $x_2 > x_0$. Combining the above facts we obtain Figure 1, the graph of $F(x_1, x_2)$ as a function of $x_2$.



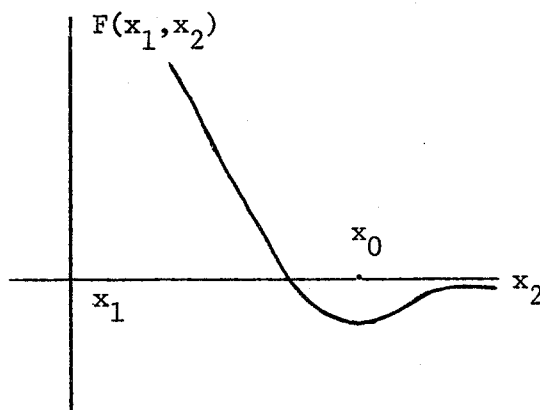Figure 1

Hence the equation has a unique solution. Since $F'_{x_2}(x_1, x_2) < 0$ in the neighborhood of $(x_1, h(x_1))$ and

$$F'_{x_1}(x_1, x_2) = \tfrac{1}{4}(x_2 - x_1)\phi(\tfrac{1}{2}(x_1 + x_2)) > 0 \ ,$$

we see $dx_2/dx_1 > 0$ in the neighborhood of $(x_1, h(x_1))$ for all $x_1 \geq 0$. Q.E.D.

14

Remark 1.  In particular, when $x_1 = 0$ the equation (4.2)' reduces to

$$(4.3) \qquad \phi(\tfrac{1}{2} x) = x(1 - \Phi(\tfrac{1}{2} x)) \; .$$

It has a unique solution and the solution is 1.224014 by the bisection procedure or Newton procedure.  It is the solution for $x_1$ in the case of $m = 3$, i.e., $x_{31} = 1.224014$.

Remark 2.  It can be verified that when $x_1$ is from $0$ to $\infty$, $x_2 = h(x_1)$ is from $x_{31}$ to $\infty$.

Lemma 4.  If $0 < y < \sqrt{(8/3) \log 2}$

$$(4.4) \qquad \tfrac{1}{4} y \phi(\tfrac{1}{2} y) < \Phi(y) - \Phi(\tfrac{1}{2} y) \; .$$

Proof.  Let

$$f(y) = \tfrac{1}{4} y \phi(\tfrac{1}{2} y) - \Phi(y) + \Phi(\tfrac{1}{2} y) \; .$$

As

$$\tfrac{1}{2} y \phi(y) < \frac{1}{\sqrt{2\pi}} \int_{\tfrac{1}{2} y}^{y} e^{-\tfrac{1}{2} t^2} dt < \tfrac{1}{2} y \phi(\tfrac{1}{2} y) \; ,$$

we have

$$f(y) < \tfrac{1}{4} y \phi(\tfrac{1}{2} y) - \tfrac{1}{2} y \phi(y) = \tfrac{1}{4} y (\phi(\tfrac{1}{2} y) - 2\phi(y))$$

$$= \tfrac{1}{4} y \phi(\tfrac{1}{2} y) (1 - 2e^{-(3/8)y^2}) < 0 \; ,$$

if $y^2 < (8/3) \log 2$.  Q.E.D.

Theorem 3. For any given $x_1 > 0$, the equation

(4.5)
$$x_1 M(\tfrac{1}{2} x_1, \tfrac{1}{2}(x_1 + x_2)) = 1$$

or

(4.5)'
$$\phi(\tfrac{1}{2} x_1) - \phi(\tfrac{1}{2}(x_1 + x_2)) = x_1 [\Phi(\tfrac{1}{2}(x_1 + x_2)) - \Phi(\tfrac{1}{2} x_1)]$$

has a unique solution $x_2 = t_2(x_1)$ if and only if $x_1 < x_{31} (= 1.224014)$. And $t_2(x_1)$ is a strictly increasing function.

Proof. Let

$$F(x_1, x_2) = 1 - x_1 M(\tfrac{1}{2} x_1, \tfrac{1}{2}(x_1 + x_2)) .$$

From Lemma 2 we have

$$F'_{x_2}(x_1, x_2) = -\tfrac{1}{2} x_1 M'_y(\tfrac{1}{2} x_1, y)\Big|_{y = \tfrac{1}{2}(x_1 + x_2)} > 0$$

and

$$F(x_1, x_1) = 1 - x_1 M(\tfrac{1}{2} x_1, x_1) < 0 .$$

Thus the equation (4.5) has a unique solution if and only if $F(x_1, \infty) > 0$. Noting $F(x_1, \infty) = 1 - x_1 M(\tfrac{1}{2} x_1, \infty) = 1 - x_1 M(\tfrac{1}{2} x_1)$, from the property of the equation (4.3), $F(x_1, \infty) > 0$ if and only if $x_1 < x_{31}$ (cf. Figure 1.).

In order to prove $t_2(x_1)$ is monotonic, we start at (4.5)'. Let

$$G(x_1, x_2) = \phi(\tfrac{1}{2} x_1) - \phi(\tfrac{1}{2}(x_1 + x_2)) - x_1 [\Phi(\tfrac{1}{2}(x_1 + x_2)) - \Phi(\tfrac{1}{2} x_1)] .$$

We have

16

$$G'_{x_2}(x_1,x_2) = \tfrac{1}{4}(x_2-x_1)\,\phi(\tfrac{1}{2}(x_1+x_2)) > 0,$$

$$G'_{x_1}(x_1,x_2) = \tfrac{1}{4}x_1\phi(\tfrac{1}{2}x_1) + \tfrac{1}{4}(x_2-x_1)\phi(\tfrac{1}{2}(x_1+x_2)) - \Phi(\tfrac{1}{2}(x_1+x_2)) + \Phi(\tfrac{1}{2}x_1) \equiv H(x_1,x_2),$$

say. Then $dx_2/dx_1 > 0$ if $H(x_1,x_2) < 0$ for $0 < x_1 \leq x_2$ and $x_1 < x_{31}$.
By Lemma 4, for any $0 < x_1 < x_{31}(< (8/3)\ln 2)$

$$H(x_1,x_1) = \tfrac{1}{4}x_1\,\phi(\tfrac{1}{2}x_1) - [\Phi(x_1) - \Phi(\tfrac{1}{2}x_1)] < 0,$$

$$H'_{x_2}(x_1,x_2) = -\frac{1}{16}\phi(\tfrac{1}{2}(x_1+x_2))(4+x_2^2-x_1^2) < 0.$$

The theorem follows.    Q.E.D.

It can be shown that $x_2 = t_2(x_1)$ is from zero to $\infty$ as $x_1$ goes from zero to $x_{31}$.

In the case of $m = 4(k = 2)$, the system (2.5) reduces to

$$(4.1)' \qquad \phi(0) - \phi(\tfrac{1}{2}(x_1+x_2)) = x_1[\Phi(\tfrac{1}{2}(x_1+x_2)) - \tfrac{1}{2}],$$

$$(4.2)' \qquad \phi(\tfrac{1}{2}(x_1+x_2)) = x_2[1-\Phi(\tfrac{1}{2}(x_1+x_2))].$$

From any given $x_1 > 0$ we can obtain $x_2 = g_2(x_1)$ (if $x_1 < x_{21}$) from (4.1)' and $x_2^* = h(x_1)$ from (4.2)', respectively. Figure 2 is the graph of $x_2$ and $x_2^*$. By computation (in detail later), we find $x_{41} = 0.452781$ and $x_{42} = 1.510437$.
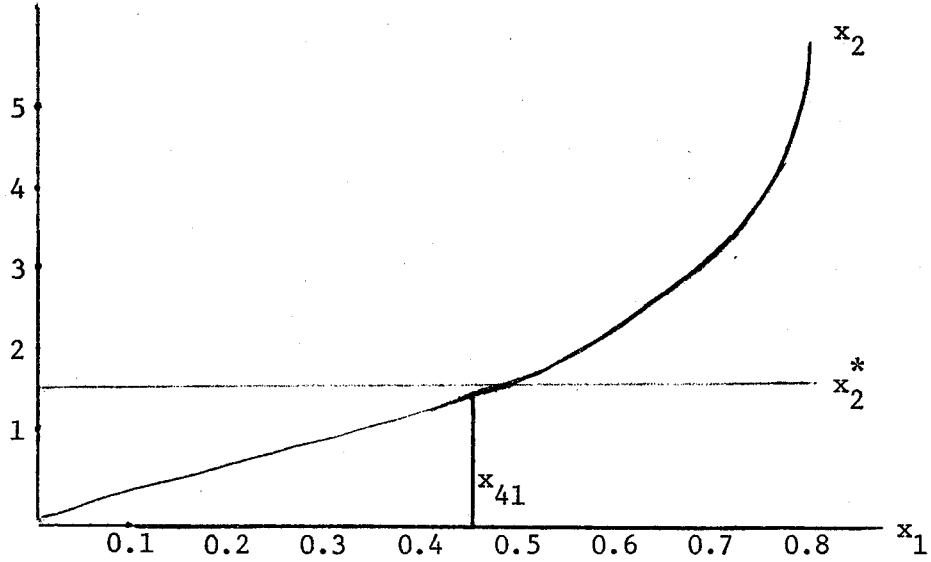
Figure 2

In the case of $m = 2k(k > 2)$, we need to solve the system of equations (2.5). From Theorem 1, for any $0 < x_1 < x_{21}$ there exists a unique solution $x_2 = g_2(x_1)$ by the first equation of (2.5). We wish to obtain $x_3 = g_3(x_1)$ from the second equation of (2.5) based on the $x_1$ and $x_2 = g_2(x_1)$.

<u>Lemma 5</u>. For any given $(x_1, g_2(x_1))$, the equation

$$(4.6) \qquad x_2 M(\tfrac{1}{2}(x_1 + x_2),\ \tfrac{1}{2}(x_2 + x_3)) = 1$$

or

$$(4.6)' \qquad \phi(\tfrac{1}{2}(x_1 + x_2)) - \phi(\tfrac{1}{2}(x_2 + x_3)) = x_2[\Phi(\tfrac{1}{2}(x_2 + x_3)) - \Phi(\tfrac{1}{2}(x_1 + x_2))]$$

has a unique solution $x_3 = g_3(x_1)$ if and only if $x_1 < x_{41}$.

18

Proof. Let

$$F(x_1,x_2,x_3) = 1 - x_2 M(\tfrac{1}{2}(x_1+x_2), \tfrac{1}{2}(x_2+x_3)) , \quad x_3 \geq x_2 \geq x_1 .$$

By using Lemma 2 we have

$$F'_{x_3}(x_1,x_2,x_3) = -\tfrac{1}{2} x_2 M'_y(\tfrac{1}{2}(x_1+x_2),y)\Big|_{y=\tfrac{1}{2}(x_2+x_3)} > 0$$

and

$$F(x_1,x_2,x_2) = 1 - x_2 M(\tfrac{1}{2}(x_1+x_2),x_2) < 0 .$$

Thus the equation (4.6) has a unique solution if and only if $F(x_1,x_2,\infty) > 0$. As

$$F(x_1,x_2,\infty) = 1 - x_2 M(\tfrac{1}{2}(x_1+x_2))$$

it is the function corresponding to the equation (4.2). From Figure 2 and thy proof of Theorem 2, $F(x_1,x_2,\infty) > 0$ if and only if $x_1 < x_{41}$. Q.E.D.

Similarly, for given $x_1$ we have in turn obtained $x_2 = g_2(x_1),\ldots,x_i = g_i(x_1)$ from the 1st, 2nd,...,(i-1)th equation of (2.5), then the equation

$$(4.7) \qquad x_i M(\tfrac{1}{2}(x_{i-1}+x_i), \tfrac{1}{2}(x_i+x_{i+1})) = 1$$

has a unique solution $x_{i+1} = g_{i+1}(x_1)$ if and only if $x_1 < x_{2i,1}$. A similar conclusion is correct for system (2.6).

By computation we obtain Figure 3 and Figure 4 for the case of $m = 2k$. Figure 3 shows us that each $x_i = g_i(x_1)$ is a strictly increasing function of $x_1$. And Figure 4 expresses the relationship between $x_{k-1}$ and $x_k$ (depending on 1) and between $x_{k-1}$ and $x_k^*$ which is obtained by solving the last equation of (2.5) for the same $x_{k-1}$.

Figure 3

Figure 4
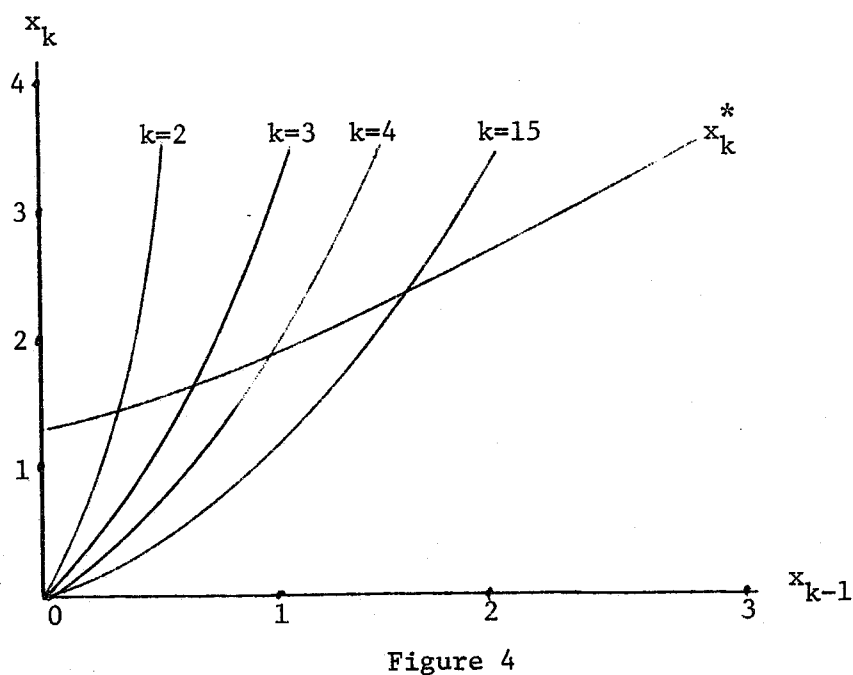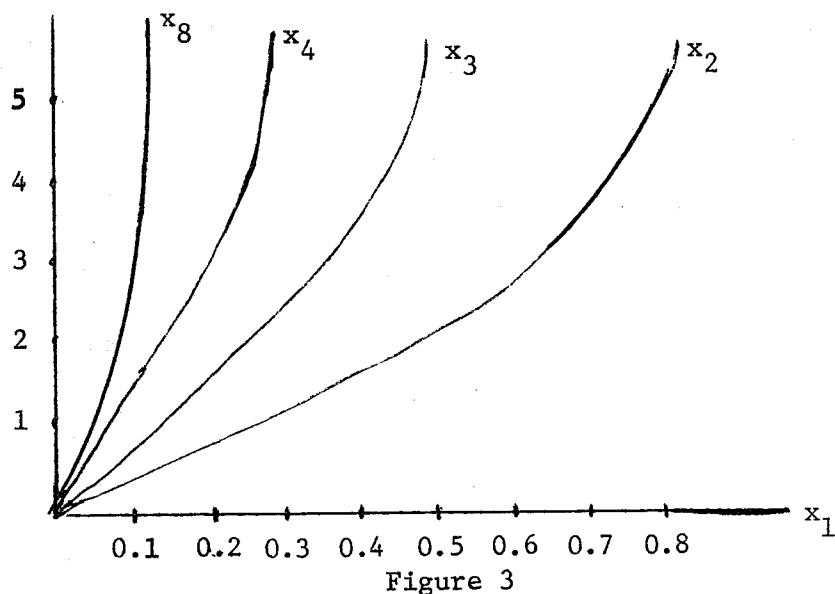
## 5. Computational Procedure and Results.

Based on the discussion in Section 4, a computational procedure is given as follows: If $m = 2k$ do the following steps:

1) Take $x_1 = (k/(k+1))x_{2(k-1),1}$ as the starting point with $x_{21} = 0.797885$ and let $LP = 0$ and $RP = x_{2(k-1),1}$, where $(LP,RP)$ is the interval in which $x_{2k,1}$ will fall.

2) For given $x_1$, in turn obtaining $x_2 = g_2(x_1), \ldots, x_k = g_k(x_1)$ from the first $(k-1)$ equations of (2.5). In other words, for the $x_{k-1}$ solve $x_k$ from the last equation of (2.5) and denote the solution as $x_k^*$.

3) Let $\varepsilon$ be a given small constant. (We take $\varepsilon = 10^{-5}$.) There are the following possible situations:

a) $|x_k - x_k^*| < \varepsilon$, then $x_1, \ldots, x_k$ are taken as the solution of (2.5), i.e., $x_{2k,1} = x_1, \ldots, x_{2k,k} = x_k$.

b) $x_k < x_k^* + \varepsilon$, Figure 4 shows that the starting $x_1$ is too small. Let $LP = x_1$, and $x_1 = \frac{1}{2}(LP+RP)$, and return to 2).

c) $x_k > x_k^* - \varepsilon$, $x_1$ is too large. Let $RP = x_1$ and $x_1 = \frac{1}{2}(LP+RP)$ and return to 2).

Obviously, the above procedure converges to the solution, because the length of the interval $(LP,RP)$ reduces to half of the original one after each repeat. There is a similar procedure for the case of $m = 2k+1$.

Tables 1 and 2 list all of $x_{m,j}$ for $m$ being even and $m$ being odd, $m \leq 31$, respectively. If we want to classify the population into $m$ subclasses, the cut-off points are $\pm(x_{m,j}+x_{m,j+1})/2$, $j = 1, 2, \ldots, [m/2]-1$, where $[x]$ denotes the largest integer which is less than or equal to $x$.

21

When  m  is even, zero is a cut-off point.  Max (1960) listed  $\{x_{mj}\}$
and cut-off points up to  $m \leq 36$  but Tables 1 and 2 have more decimal
places than those given by Max.  The first column in Table 1 and Table 2
lists  values of  (1 - loss functions)%  which give us the information
for determining the value of  m.

Table 3 and Table 4 give probabilities being represented by  $\{x_{m,j}\}$
for  m  being even  and  m  being odd, respectively.  If  $m = 2k+1$,
$x_{m,0} = 0$,  and we need list it; but the corresponding probability
should be listed.  Hence the forms of Table 2 and Table 4 are a little
different.

Let us look at a use of the Tables.  Suppose we want to design
clothes for a group of people and know that height in this group is
distributed according to  $N(\mu, \sigma^2)$  with  $\mu = 170$cm  and  $\sigma = 10$cm.
Let  m  be the number of sizes that we wish to produce, that is we
wish to determine representative heights for  m  models.  If  $m = 7$,
we find from Table 2 that the representative points are  $x_{70} = 0$,
$x_{71} = 0.560607$,  $x_{72} = 1.188219$  and  $x_{73} = 2.033827$.  Therefore, the
heights of  $m = 7$  models are  $\mu \pm x_{7i}\sigma$,  $i = 1,2,3$,  i.e. 149.7, 158.1,
164.6, 170.0, 175.6, 181.9 and 190.7.  The first column of Table 2
shows us that use of these 7 height categories instead of the continuum
represents a loss of information equal to 4.4% as measured by our mini-
mized square loss function (1.1).  If we partition the group into 7 sub-
groups, the cut-off points are 153.9 ((149.7+158.1)/2), 161.4, 167.3,
172.8, 178.8, and 186.3.  The relative frequencies associated with these sub-
groups are 5.36%, 13.74%, 19.87%, 22.08%, 19.87%, 13.74% and 5.36% by Table 4.
If we assume  $m = 6$,  then from Table 1 and Table 3, the representative
points are 151.1, 160.0, 166.8, 173.2, 180.0 and 188.9, the cut-off points

22

are 155.6, 163.4, 170.0, 176.6 and 184.5, with corresponding relative frequencies 7.39%, 18.10%, 24.50, 24.50%, 18.01% and 7.39%; and the loss of information is 5.8%. Certainly for $m \geq 10$, the increase in information is quite negligible. It is only for the smaller $m$ values that increasing or decreasing the number of categories may be significant in the tradeoff with information gain or loss.

Acknowledgments.

Kai-Tai Fang
Stanford University and
Institute of Applied Mathematics, Academia Sinica


Shu-Dong He
Stanford University and
Huazheng Institute of Technology, Wuhan, China

## TABLE 1  X(m, j)  (m — even)

| m | 1-LOSS FU | j=1 | j=2 | j=3 | j=4 | j=5 | j=6 | j=7 |
|---|---|---|---|---|---|---|---|---|
| 2 | 63.661990 | 0.797885 | | | | | | |
| 4 | 88.252192 | 0.452781 | 1.510437 | | | | | |
| 6 | 94.204586 | 0.317726 | 1.000143 | 1.893830 | | | | |
| 8 | 96.549558 | 0.245112 | 0.756060 | 1.344048 | 2.152724 | | | |
| 10 | 97.712362 | 0.199659 | 0.609934 | 1.057975 | 1.591756 | 2.346883 | | |
| 12 | 98.373484 | 0.168469 | 0.511949 | 0.876964 | 1.286063 | 1.783980 | 2.501750 | |
| 14 | 98.785352 | 0.145751 | 0.441444 | 0.750661 | 1.085993 | 1.468236 | 1.940410 | 2.630429 |
| 16 | 99.059320 | 0.128443 | 0.388197 | 0.657018 | 0.942738 | 1.256885 | 1.619344 | 2.072074 |
| 18 | 99.250818 | 0.114824 | 0.346516 | 0.584599 | 0.834305 | 1.102760 | 1.400933 | 1.748159 |
| 20 | 99.389886 | 0.103828 | 0.312991 | 0.526833 | 0.749039 | 0.984354 | 1.239527 | 1.525192 |
| 22 | 99.494100 | 0.094761 | 0.285426 | 0.479624 | 0.680056 | 0.890055 | 1.114100 | 1.358719 |
| 24 | 99.574232 | 0.087156 | 0.262357 | 0.440290 | 0.623002 | 0.812935 | 1.013227 | 1.228196 |
| 26 | 99.637198 | 0.080687 | 0.242766 | 0.407005 | 0.574992 | 0.748576 | 0.930052 | 1.122436 |
| 28 | 99.687480 | 0.075117 | 0.225920 | 0.378458 | 0.533995 | 0.693966 | 0.860102 | 1.034604 |
| 30 | 99.728346 | 0.070272 | 0.211280 | 0.353707 | 0.498570 | 0.647011 | 0.800374 | 0.960315 |

| m | j=8 | j=9 | j=10 | j=11 | j=12 | j=13 | j=14 | j=15 |
|---|---|---|---|---|---|---|---|---|
| 16 | 2.740625 | | | | | | | |
| 18 | 2.185671 | 2.837083 | | | | | | |
| 20 | 1.860332 | 2.285642 | 2.923058 | | | | | |
| 22 | 1.634326 | 1.959668 | 2.375024 | 3.000824 | | | | |
| 24 | 1.464226 | 1.731595 | 2.048884 | 2.456031 | 3.072078 | | | |
| 26 | 1.329943 | 1.558865 | 1.819393 | 2.129989 | 2.530294 | 3.138023 | | |
| 28 | 1.220415 | 1.421689 | 1.644658 | 1.899454 | 2.204461 | 2.599053 | 3.199659 | |
| 30 | 1.128976 | 1.309258 | 1.505270 | 1.723212 | 1.973188 | 2.273515 | 2.663323 | 3.257769 |

## TABLE 2  X(m, j)  (m — odd)

| m | 1-LOSS FU | j=1 | j=2 | j=3 | j=4 | j=5 | j=6 | j=7 |
|---|---|---|---|---|---|---|---|---|
| 3 | 80.982620 | 1.224014 | | | | | | |
| 5 | 92.007208 | 0.764582 | 1.724227 | | | | | |
| 7 | 95.603322 | 0.560607 | 1.188219 | 2.033827 | | | | |
| 9 | 97.219908 | 0.443675 | 0.918894 | 1.476651 | 2.255903 | | | |
| 11 | 98.084844 | 0.367515 | 0.752492 | 1.179063 | 1.693290 | 2.428242 | | |
| 13 | 98.601734 | 0.313847 | 0.638403 | 0.987206 | 1.381766 | 1.865845 | 2.568793 | |
| 15 | 98.935282 | 0.273945 | 0.554949 | 0.851433 | 1.175366 | 1.547025 | 2.008841 | 2.687479 |
| 17 | 99.163116 | 0.243098 | 0.491099 | 0.749630 | 1.026111 | 1.331750 | 1.686128 | 2.130809 |
| 19 | 99.325586 | 0.218530 | 0.440605 | 0.670188 | 0.912228 | 1.173637 | 1.465193 | 1.806041 |
| 21 | 99.445534 | 0.198496 | 0.399639 | 0.606335 | 0.822066 | 1.051424 | 1.301035 | 1.581410 |
| 23 | 99.536706 | 0.181846 | 0.365723 | 0.553826 | 0.748709 | 0.953596 | 1.172873 | 1.412985 |
| 25 | 99.607540 | 0.167788 | 0.337169 | 0.509846 | 0.687751 | 0.873242 | 1.069389 | 1.280453 |
| 27 | 99.663686 | 0.155762 | 0.312798 | 0.472454 | 0.636232 | 0.805914 | 0.983741 | 1.172690 |
| 29 | 99.708938 | 0.145357 | 0.291748 | 0.440256 | 0.592077 | 0.748591 | 0.911489 | 1.082946 |
| 31 | 99.745920 | 0.136258 | 0.273375 | 0.412228 | 0.553784 | 0.699141 | 0.849610 | 1.006826 |

| m | j=8 | j=9 | j=10 | j=11 | j=12 | j=13 | j=14 | j=15 |
|---|---|---|---|---|---|---|---|---|
| 17 | 2.790329 | | | | | | | |
| 19 | 2.237125 | 2.881189 | | | | | | |
| 21 | 1.911404 | 2.331491 | 2.962832 | | | | | |
| 23 | 1.684276 | 2.005400 | 2.416462 | 3.037181 | | | | |
| 25 | 1.512767 | 1.776563 | 2.090355 | 2.493931 | 3.105654 | | | |
| 27 | 1.376945 | 1.602762 | 1.860301 | 2.167977 | 2.565299 | 3.169331 | | |
| 29 | 1.265876 | 1.464410 | 1.684760 | 1.937043 | 2.239604 | 2.631692 | 3.227091 | |
| 31 | 1.172913 | 1.350754 | 1.544445 | 1.760177 | 2.008046 | 2.306347 | 2.694096 | 3.285904 |

TABLE 3

| m | J=0 | J=1 | J=2 | J=3 | J=4 | J=5 | J=6 | J=7 |
|---|---|---|---|---|---|---|---|---|
| 2 | | 0.500000 | | | | | | |
| 4 | | 0.336855 | 0.163145 | | | | | |
| 6 | | 0.245031 | 0.181023 | 0.073946 | | | | |
| 8 | | 0.191669 | 0.161486 | 0.106654 | 0.040192 | | | |
| 10 | | 0.157187 | 0.140660 | 0.109548 | 0.068155 | 0.024450 | | |
| 12 | | 0.133151 | 0.123152 | 0.103965 | 0.077349 | 0.046331 | 0.016053 | |
| 14 | | 0.115467 | 0.108963 | 0.096347 | 0.078445 | 0.056624 | 0.033019 | 0.011135 |
| 16 | | 0.101920 | 0.097456 | 0.088733 | 0.076185 | 0.060509 | 0.042737 | 0.024412 |
| 18 | | 0.091214 | 0.088020 | 0.081744 | 0.072633 | 0.061078 | 0.047635 | 0.033087 |
| 20 | | 0.082545 | 0.080180 | 0.075517 | 0.068704 | 0.059974 | 0.049653 | 0.038183 |
| 22 | | 0.075382 | 0.073582 | 0.070025 | 0.064801 | 0.058058 | 0.050002 | 0.040897 |
| 24 | | 0.069364 | 0.067963 | 0.065187 | 0.061097 | 0.055789 | 0.049396 | 0.042090 |
| 26 | | 0.064239 | 0.063127 | 0.060920 | 0.057659 | 0.053409 | 0.048259 | 0.042326 |
| 28 | | 0.059822 | 0.058924 | 0.057140 | 0.054499 | 0.051045 | 0.046840 | 0.041967 |
| 30 | | 0.055976 | 0.055241 | 0.053779 | 0.051610 | 0.048766 | 0.045291 | 0.041244 |

| m | J=8 | J=9 | J=10 | J=11 | J=12 | J=13 | J=14 | J=15 |
|---|---|---|---|---|---|---|---|---|
| 16 | 0.008048 | | | | | | | |
| 18 | 0.018583 | 0.006005 | | | | | | |
| 20 | 0.026166 | 0.014483 | 0.004595 | | | | | |
| 22 | 0.031093 | 0.021065 | 0.011508 | 0.003588 | | | | |
| 24 | 0.034088 | 0.025671 | 0.017217 | 0.009290 | 0.002848 | | | |
| 26 | 0.035753 | 0.028717 | 0.021448 | 0.014253 | 0.007599 | 0.002291 | | |
| 28 | 0.036523 | 0.030627 | 0.024426 | 0.018107 | 0.011929 | 0.006286 | 0.001864 | |
| 30 | 0.036695 | 0.031727 | 0.026439 | 0.020953 | 0.015426 | 0.010076 | 0.005247 | 0.001530 |

TABLE 4

| m | J=0 | J=1 | J=2 | J=3 | J=4 | J=5 | J=6 | J=7 |
|---|---|---|---|---|---|---|---|---|
| 3 | 0.459467 | 0.270266 | | | | | | |
| 5 | 0.297755 | 0.244450 | 0.106673 | | | | | |
| 7 | 0.220756 | 0.198676 | 0.137365 | 0.053581 | | | | |
| 9 | 0.175560 | 0.164374 | 0.132345 | 0.084507 | 0.030993 | | | |
| 11 | 0.145796 | 0.139364 | 0.120661 | 0.091605 | 0.055818 | 0.019654 | | |
| 13 | 0.124695 | 0.120660 | 0.108828 | 0.090057 | 0.065906 | 0.038911 | 0.013292 | |
| 15 | 0.108947 | 0.106253 | 0.098305 | 0.085535 | 0.068711 | 0.049022 | 0.028276 | 0.009424 |
| 17 | 0.096744 | 0.094855 | 0.089264 | 0.080204 | 0.068094 | 0.053559 | 0.037495 | 0.021230 |
| 19 | 0.087007 | 0.085633 | 0.081552 | 0.074901 | 0.065922 | 0.054977 | 0.042558 | 0.029353 |
| 21 | 0.079058 | 0.078027 | 0.074958 | 0.069935 | 0.063107 | 0.054695 | 0.044993 | 0.034393 |
| 23 | 0.072446 | 0.071652 | 0.069286 | 0.065402 | 0.060095 | 0.053505 | 0.045822 | 0.037285 |
| 25 | 0.066859 | 0.066235 | 0.064373 | 0.061309 | 0.057105 | 0.051854 | 0.045684 | 0.038750 |
| 27 | 0.062077 | 0.061578 | 0.060086 | 0.057626 | 0.054240 | 0.049993 | 0.044972 | 0.039283 |
| 29 | 0.057938 | 0.057532 | 0.056318 | 0.054313 | 0.051546 | 0.048065 | 0.043929 | 0.039215 |
| 31 | 0.054317 | 0.053984 | 0.052984 | 0.051328 | 0.049040 | 0.046152 | 0.042708 | 0.038763 |

| m | J=8 | J=9 | J=10 | J=11 | J=12 | J=13 | J=14 | J=15 |
|---|---|---|---|---|---|---|---|---|
| 17 | 0.006927 | | | | | | | |
| 19 | 0.016363 | 0.005238 | | | | | | |
| 21 | 0.023430 | 0.012882 | 0.004051 | | | | | |
| 23 | 0.028207 | 0.019012 | 0.010321 | 0.003190 | | | | |
| 25 | 0.031248 | 0.023432 | 0.015642 | 0.008389 | 0.002550 | | | |
| 27 | 0.033057 | 0.026455 | 0.019683 | 0.013022 | 0.006903 | 0.002064 | | |
| 29 | 0.034013 | 0.028430 | 0.022600 | 0.016695 | 0.010952 | 0.005737 | 0.001687 | |
| 31 | 0.034382 | 0.029642 | 0.024631 | 0.019462 | 0.014281 | 0.009289 | 0.004808 | 0.001390 |

# References

Anderberg, M. R. (1973). Cluster Analysis for Applications, Academic Press.

Bofinger, E. (1970). Maximizing the correlation of grouped observations, J. American Statistical Association, 65, 1632-1638.

Cox, D. R. (1957). Note on grouping, J. American Statistical Association, 52, 543-547.

Fang, K. T. (1976). Application of the theory of the conditional distribution for the standardization of clothes, Acta Mathematicae Applagatae Sinica, 2, 62-74. (in Chinese)

Fang, K. T., and Wu, C.Y. (1979). The extreme value problem of some probability functions, Acta Mathematicae Applagatae Sinica, 2, 132-148. (in Chinese).

Feller, W. (1968). An Introduction to Probability Theory and Its Applications, Vol. I (third edition), John Wiley & Sons, Inc.

Max, J. (1960). Quantizing for minimum distribution, Institute of Radio Engineers Transactions on Information Theory, 7-12.

Sitgreaves, R. (1961). Optimal test design in a special situation, Chapter 3, Studies in Item Analysis and Prediction, edited by Herbert Solomon, Stanford University Press, Stanford, California.

Zador, P. (1963). Development and evaluation of procedures for quantizing multivariate distributions, Technical Report No. 92, Contract Nonr-225(52)(NR-342-022), Department of Statistics, Stanford University, Stanford, California.

Zador, P. (1982). Asymptotic quantization error of continuous signals and the quantization dimension. IEEE Transactions on Information Theory, Part 1, March 1982, 139-148.

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER 5 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE *(and Subtitle)* The Problem of Selecting a Given Number of Representative Points in a Normal Population and a Generalized Mills' Ratio | | 5. TYPE OF REPORT & PERIOD COVERED Technical Report |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR*(s)* Kai-Tai Fang Shu-Dong He | | 8. CONTRACT OR GRANT NUMBER*(s)* DAAG 29-82-K-0156 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-19065-M |
| 11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park. NC 27709 | | 12. REPORT DATE October 1982 |
| | | 13. NUMBER OF PAGES 26 |
| 14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* | | 15. SECURITY CLASS. *(of this report)* UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. ~~DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)~~

Also issued as Technical Report No. 327 under Office of Naval Research Contract No. N00014-76-C-0475, Department of Statistics, Stanford University.

18. SUPPLEMENTARY NOTES
THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Bisection procedure, generalized Mills' ratio, Quantization, Newton procedure, representative points.

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

PLEASE SEE REVERSE SIDE.

20.

In a known normal population, one desires to select a given number of representative points that retain as much information about the population as possible. To obtain these points, the solution of two systems of equations is required. Some properties of these equations are discussed. As a natural consequence a generalized Mills' ratio is defined. Some basic properties of the generalized Mills' ratio for studying the above equation are listed. A computational procedure is given to obtain these points and tables of them and corresponding probabilities are given for $m \leq 31$, where $m$ is the number of points.

#5ARO/#3270NR